ID 300739 (available ainsi que ID 300589)
ID 300940 ID 300602

Citation for PHFR 000129 EP-p.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: H04N 7/26 — A1

(11) International Publication Number: WO 00/01154

(43) International Publication Date: 6 January 2000 (06.01.00)

(21) International Application Number: PCT/US99/14306

(22) International Filing Date: 24 June 1999 (24.06.99)

(30) Priority Data:
60/090,845 26 June 1998 (26.06.98) US

(71) Applicant (*for all designated States except US*): GENERAL INSTRUMENT CORPORATION [US/US]; 101 Tournament Drive, Horsham, PA 19044 (US).

(72) Inventor; and

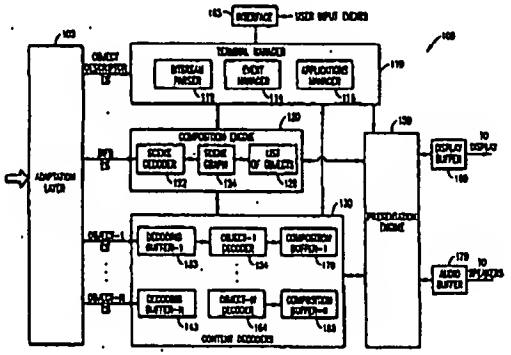
(75) Inventor/Applicant (*for US only*): RAJAN, Ganesh [IN/US]; 13659 Tiverton Road, San Diego, CA 92130 (US).

(74) Agent: LIPSITZ, Barry, R.; Building 8, 755 Main Street, Monroe, CT 06468 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: TERMINAL FOR COMPOSING AND PRESENTING MPEG-4 VIDEO PROGRAMS



(57) Abstract

A method and apparatus for composing and presenting multimedia programs using the MPEG-4 standard at a multimedia terminal (100). A composition engine (120) maintains and updates a scene graph (124) of the current objects, including their relative position in a scene and their characteristics, and provides a corresponding list of objects (126) to be displayed to a presentation engine (150). In response, the presentation engine begins to retrieve the corresponding decoded object data that is stored in respective composition buffers (176, . . . 186). The presentation engine assembles the decoded objects to provide a scene for presentation on output devices such as a video monitor (240) and speakers (242), or for storage. A terminal manager (110) receives user commands and causes the composition engine to update the scene graph and list of objects accordingly. The terminal manager also forwards the information contained in the object descriptors to a scene decoder (122) at the composition engine.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

TERMINAL FOR COMPOSING AND PRESENTING MPEG-4 VIDEO PROGRAMS

BACKGROUND OF THE INVENTION

This application claims the benefit of U.S. Provisional Application No. 60/090,845, filed June 26, 1998.

The present invention relates to a method and apparatus for composing and presenting multimedia video programs using the MPEG-4 (Motion Picture Experts Group) standard. More particularly, the present invention provides an architecture wherein the composition of a multimedia scene and its presentation are processed by two different entities, namely a "composition engine" and a "presentation engine."

The MPEG-4 communications standard is described, e.g., in ISO/IEC 14496-1 (1999): Information Technology - Very Low Bit Rate Audio-Visual Coding - Part 1" Systems; ISO/IEC JTC1/SC29/WG11, MPEG-4 Video Verification Model Version 7.0 (February 1997); and ISO/IEC JTC1/SC29/WG11 N2725, MPEG-4 Overview (March 1999/Seoul, South Korea).

The MPEG-4 communication standard allows a user to interact with video and audio objects within a scene, whether they are from conventional sources, such as moving video, or from synthetic (computer-generated) sources. The user can modify scenes by

deleting, adding or repositioning objects, or changing the characteristics of the objects, such as size, color, and shape, for example.

The term "multimedia object" is used to encompass audio and/or video objects.

The objects can exist independantly, or be joined with other objects in a scene in a grouping known as a "composition". Visual objects in a scene are given a position in two- or three-dimensional space, while audio objects can be placed in a sound space.

MPEG-4 uses a syntax structure known as Binary Format for Scenes (BIFS) to describe and dynamically change a scene. The necessary composition information forms the scene description, which is coded and transmitted together with the media objects. BIFS is based on VRML (the Virtual Reality Modeling Language). Moreover, to facilitate the development of authoring, manipulation and interaction tools, scene descriptions are coded independently from streams related to primitive media objects.

BIFS commands can add or delete objects from a scene, for example, or change the visual or acoustic properties of objects. BIFS commands also define, update, and position the objects. For example, a visual property such as the color or size of an object can be changed, or the object can be animated.

The objects are placed in elementary streams (ESs) for transmission, e.g., from a headend to a

decoder population in a broadband communication network, such as a cable or satellite television network, or from a server to a client PC in a point-to-point Internet communication session. Each object is carried in one or more associated ESs. A scaleable object may have two ESs for example, while a non-scaleable object has one ES. Data that describes a scene, including the BIFS data, is carried in its own ES.

Furthermore, MPEG-4 defines the structure for an object descriptor (OD) that informs the receiving system which ESs are associated with which objects in the received scene. ODs contain elementary stream descriptors (ESDs) to inform the system which decoders are needed to decode a stream. ODs are carried in their own ESs and can be added or deleted dynamically as a scene changes.

A synchronization layer, at the sending terminal, fragments the individual ESs into packets, and adds timing information to the payload of these packets. The packets are then passed to the transport layer and subsequently to the network layer, for communication to one or more receiving terminals.

At the receiving terminal, the synchronization layer parses the received packets, assembles the individual ESs required by the scene, and makes them available to one or more of the appropriate decoders.

The decoder obtains timing information from an encoder clock, and time stamps of the incoming

streams, including decode time stamps and composition time stamps.

MPEG-4 does not define a specific transport mechanism, and it is expected that the MPEG-2 transport stream, asynchronous transfer mode, or the Internet's Real-time Transfer Protocol (RTP) are appropriate choices.

The MPEG-4 tool "FlexMux" avoids the need for a separate channel for each data stream. Another tool (Digital Media Interface Format - DMIF) provides a common interface for connecting to varying sources, including broadcast channels, interactive sessions, and local storage media, based on quality of services (QoS) factors.

Moreover, MPEG-4 allows arbitrary visual shapes to be described using either binary shape encoding, which is suitable for low bit rate environments, or grey scale encoding, which is suitable for higher quality content.

However, MPEG-4 does not specify how shapes and audio objects are to be extracted and prepared for display or play, respectively.

Accordingly, it would be desirable to provide a general architecture for a decoding system that is capable of receiving and presenting programs conforming to the MPEG-4 standard.

The terminal should be capable of composing and presenting MPEG-4 programs.

The composition of a multimedia scene and its presentation should be separated into two entities,

i.e., a composition engine and a presentation engine.

The scene compoeition data, received in the BIFS format, should be decoded and transleted into a scene graph in the composition engine.

The system ehould incorporate updetes to a scene, received via the BIFS stream or vie local interection, into the scene graph in the composition engine.

The compoeition engine should make evailable a liet of multimedie objects (including displayable and/or eudible objecte) to the presentation engine for presentetion, sufficiently prior to each preeentation instant.

The presentetion engine should read the objects to be presented from the list, retrieve the objects from content decodere, and render the objecte into sppropriate buffers (e.g., displey and audio buffers).

The compoeition and presentation of content should preferably be performed independently so that the prsentation engine does not have to wait for the composition engine to finish ite tesks before the presentation engine accessee the presentable objecte.

The terminal should be suiteble for use with both broedband communication networks, such ae cable and setellite television networks, as wsll as computer networks, such ee ths Internet.

The terminal should elso be responsive to user inputs.

The system should be independsnt of the underlying transport, network and link protocols.

The present invention providea a system having the above and other advantages.

SUMMARY OF THE INVENTION

The present invention relates to a method and apparatus for composing and presenting multimedia video programs using the MPEG-4 standard.

A multimedia terminal includes a terminal manager, a composition engine, content decoders, and a presentation engine. The composition engine maintains and updates a scene graph of the current objects, including their relative position in a scene and their characteristics, to provide a list of objects to be displayed or played to the presentation engine. The list of objects is used by the presentation engine to retrieve the decoded object data that is stored in respective composition buffers of content decoders.

The presentation engine assembles the decoded objects according to the list to provide a scene for presentation, e.g., display and playing on a display device and audio device, respectively, or storage on a storage medium.

The terminal manager receives user commands and causes the composition engine to update the scene graph and list of objects in response thereto.

Moreover, the composition and the presentation of the content are preferably performed independently (i.e., with separate control threads).

Advantageously, the separate control threads allow the presentation engine to begin retrieving the corresponding decoded multimedia objects while the composition engine recovers additional scene

description information from the bitstream and/or processes additional object descriptor information provided to it.

A composition engine and a presentation engine should have the ability to communicate with each other via interfaces that facilitate the passing of messages and other data between themselves.

A terminal for receiving and processing a multimedia data bitstream, and a corresponding method are disclosed.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a general architecture for a multimedia receiver terminal capable of receiving and presenting programs conforming to the MPEG-4 standard in accordance with the present invention.

FIG. 2 illustrates the presentation process in the terminal architecture of FIG. 1 in accordance with the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a method and apparatus for composing and presenting multimedia video programs using the MPEG-4 standard.

FIG. 1 illustrates a general architecture for a multimedia receiver terminal capable of receiving and presenting programs conforming to the MPEG-4 standard in accordance with the present invention.

According to the MPEG-4 Systems standard, the scene description information is coded into a binary format known as BIFS (Binary Format for Scene). This BIFS data is packetized and multiplexed at a transmission site, such as a cable and or satellite television headend, or a server in a computer network, before being sent over a communication channel to a terminal 100. The data may be sent to a single terminal or to a terminal population. Moreover, the data may be sent via an open-access network or via a subscriber network.

The scene description information describes the logical structure of a scene, and indicates how objects are grouped together. Specifically, an MPEG-4 scene follows a hierarchical structure, which can be represented as a directed acyclic (tree) graph, where each node or a group of nodes, of the graph, represents a media object. The tree structure is not necessarily static, since node attributes (e.g., positioning parameters) can be changed while nodes can be added, replaced, or removed.

The scene description information cen also indicata how objacts ara positionad in spsca and time. In the MPEG-4 model, objects hsve both spatial and temporal charecteristica. Esch object has a local coordinata systam in which tha object has e fixed spetial-temporal location end scele. Objacts are positioned in a scena by specifying a coordinate transformation from the object's local coordinate system into a global coordinate system defined by ona more parent scene description nodes in the trae.

Tha scena description information can also indicate attributa value sslection. Individual media objects and scana description nodes exposa a set of parametars to a composition layar through which part of their behevior can be controllad. Examplas includa the pitch of a aound, the color for a synthetic object, activation or deactivation of anhancament information for scaleabla coding, and so forth.

The scene description information can also indicate other transforms on media objects. Tha scene description structura and noda ssmantics ere hsavily influenced by VRML, including its event modal. This providea MPEG-4 with en extensive sst of acane construction operators, including grsphics primitivas that cen bs used to construct sophisticstad scenes.

The "TrensMux" (Transport Multiplexing) layer of MPEG-4 modela the layer that offers transport services matching the raquested QoS. Only the

interface to this layer is specified by MPEG-4. The concreta mapping of the data packets snd control signaling may be performed using any desired transport protocol. Any suitable existing transport protocol stack, such as Real-time Transfsr Protocol (RTP)/ Uaer Dstagram Protocol (UDP)/ Internet protocol (IP), ATM Adaptation Leyer (AAL5)/ Asynchronous Transfer Mode (ATM), or MPEG-2's Transport Stream over a suitable link layar may bscome a specific TransMux instsnce. The choica is left to the end user/service provider, and allows MPEG-4 to ba usad in a wide variaty of oparational environmenta.

In the prassnt axemple, it is aasumed for illuatration only, that an ATM adeptation Layer 105 ia used for transport.

Tha multiplexed packatized streams are receivad at an input of tha multimedie tarminal 100. The various descriptors, starting with the ObjectDescriptor, ere parsed from an object descriptor ES, e.g., at a parser 112. The alementary stream descriptor (ESDescriptor), conteined within the first object descriptor (called tha Initisl ObjectDescriptor), contains s pointer locating tha Scans Deacription stream (BIFS stresm) from among the incoming multiplexed streams. In a broadcast scanario, the BIFS stream is locatsd from smong tha incoming multiplexad streams. For Internat-type scenarios, wherein there is a guarsntaed back chsnnel connection from the MPEG-4 tarminal to the underlying natwork, the BIFS stream may be retrievsd

from a remote server. The information about the various elementary streams are contained in the ObjectDescriptors and its associated descriptors. For details, see ISO/IEC CD 14496-1: Information Technology - Very low bit rate audio-visual coding - Part 1: Systems (Committee Draft of MPEG-4 Systems), incorporated herein by reference.

The parser 112, which is a general bitstream parser for the parsing of the various descriptors, is incorporated within a terminal manager 110.

The BIFS bitstream containing the scene description information is received at the BIFS Scene Decoder 122, which is shown as a component of a Composition Engine 120. The coded elementary content streams (comprising video, audio, graphics, text, etc.) are routed to their respective decoders according to the information contained in the received descriptors. The decoders for the elementary content or object streams have been grouped within a box 130 labeled "Content Decoders".

For example, an object-1 elementary stream (ES) is routed to an input decoding buffer-1 122, while an object-N ES is routed to a decoding buffer-N 132. The respective objects are decoded, e.g., at object-1 decoder 124, . . . , object-N decoder 134, and provided to respective output, composition buffers, e.g., composition buffer-1 126, . . . , composition buffer-N 136. The decoding may be scheduled based on Decode Time Stamp (DTS) information.

Note that it is possible for the data from two or more decoding buffers to be associated with one decoder, e.g., for scalable objects.

The composition engine 120 performs a variety of functions. Specifically, when a received elementary stream is a BIFS stream, the composition engine 120 creates and/or updates a scene graph at a scene graph function 124 using the output of the BIFS scene decoder 122. The scene graph provides complete information on the composition of a scene, including the types of objects present and the relative position of the objects. For example, a scene graph may indicate that a scene includes one or more persons and a synthetic, computer-generated 2-D background, and the positions of the persons in the scene.

When a received elementary stream is a BIFSAnimation stream, the appropriate spatial-temporal attributes of the components of the scene graph are updated at the scene graph function 124. Thus, the composition engine 120 maintains the status of the scene graph and its components.

From the scene graph function 124, the composition engine 120 creates a list of video objects 126 to be displayed by a presentation engine 150, and a list of audible objects to be played by the Presentation Engine 150. For generality, both video and audio objects are referred to herein as being "displayed" or "presented" on an appropriate output device. For example, video objects can be presented on a video screen, such as a television

screen or computer monitor, while audio objects can be presented via speakers. Of course, the objects can also be stored on a recording device, such as a computer's hard drive, or a digital video disc, without a user actually viewing or listening to them. The presentation engine thus provides the objects in a state in which they can be presented to some final output device, either for immediate viewing/listening and/or storage for subsequent use.

Moreover, the term "list" will be used herein to indicate any type of listing regardless of the specific implementation. For example, the list may be provided as a single list for all objects, or separate lists may be provided for different object types (e.g., video or audio), or more than one list may be provided for each object type. The list of objects is a simplified version of the scene graph information. It is only important for the presentation engine 150 to be able to use the list to recognize the objects and route them to appropriate underlying rendering engines.

The multimedia scene that is presented can include a single, still video frame or a sequence of video frames.

The composition engine 120 manages the list, and is typically the only entity that is allowed to explicitly modify the entries in the list.

Some of the presentable objects may be available in the composition buffers 126, . . . , 136 in a decoded format. If so, this is indicated

in the description of the objects in the list of objects 126.

The composition engine 120 makes the list available to the presentation engine 150 in a timely manner so that the presentation engine 150 can present the scene at the desired time instants, according to the desired presentation rate specified for the program. The presentation engine 150 presents a scene by retrieving the decoded objects from the buffers 126, . . . , 136 and providing the decoded video objects to a display buffer 160, and by providing the decoded audio objects to an audio buffer 170. The objects are subsequently presented on a display device and speakers, respectively, and/or stored at a recording device. The presentation engine 150 retrieves the decoded objects at preset presentation rates using known time stamp techniques, such as Composition Time Stamps (CTSs).

The composition engine 120 also provides the scene graph information from the scene graph function 124 to the presentation engine 150. However, the provision of the simplified list of objects allows the presentation engine to begin retrieving the decoded objects.

The composition engine 120 thus manages the scene graph. It updates the attributes of the objects in the scene graph based on factors that include a user interaction or specification, a pre-specified spatio-temporal behavior of the objects in the scene graph, which is a part of the scene graph

iteelf; and commande received on tha BIFS atream, such as BIFS updates or BIFSAnimation commands.

The compoaition engina 120 ia alao responsible for the management of the decoding buffars 122, . . . , 132 and the composition buffers 126, . . ., 136 allocatad for thia particular application by the terminal 100. For axample, the composition angine 120 ensures that thase buffare do not ovarflow or underflow. The composition engine 120 can also implement buffar control stratagies, e.g., in accordance with the MPEG-4 conformance specificationa.

The terminal manager 110 includes an avent managar 114, an applications managar 116 and a clock 118.

Multimedia applications may resida on tha terminal manager 110 aa designated by an applicationa manager 116. For exampla, thase applicationa may be include user-friandly softwara run on a PC that allowa a ueer to manipulata the objects in a acane.

Tha tarminal manager 110 managaa communicationa with tha external world through appropriate interfacae. For axample, an avent managar 114, such aa an example interface 165 which ia responsive to uaor input evante, ia responaibla for monitoring user intarfacas, and detecting the relatad aventa. Uaar input evanta include, e.g., mouse movementa and clicka, keypad clicka, joyetick movamenta, or signals from othar input devicea.

The terminal managar 110 passas tha usar input

avanta to the compoaition engina 120 for appropriate handling. For exampla, a usar may antar commands to re-position or change tha attributas of cartain objects within tha scane graph.

User interface evants may not be proceasad in aome caese, a.g., for a purely broadcast program with no interactive content.

The terminal functiona of FIG. 1 can be implamented uaing any known hardwara, firmware and/or software. Moreovar, tha varioua functional blocks ahown naed not be indapandent but can shara common hardware, firmware and/or aoftware. For example, the parser 112 can be provided outaida the terminal managar 110, a.g., in the compoaition engine 120.

Note that the content decodera 130 and compoaition angine 120 run indapendantly of each other in the sense that thair saparate control threads (a.g., control cyclea or loops) do not affact aach other. Advantagsouely, by separating tha compoaition and praeantation threads, tha presantation engina doae not have to wait for tha composition engine to finish its tasks (e.g., such aa recovaring additional acena dascription information or procaaeing object descriptors) befora the praaantation angina accassas (e.g., begina to ratrieve) the praaantable objacts from the buffera 126, . . . , 136. Thus, the presantation engine 150 runs in ita own thread and praaente the objacte at ite desired presentation rate, regardlesa of whether

the compoeition engine 120 hae finished its taske or not.

The elementary stream decoders 124, . . . , 134 also run in their individuel control threeds independent of tha presantetion and compoeition engines. Synchronization between the decoding and the composition can be achieved using conventional time stamp dete, such ee DTS, CTS and PTS dete ee they are known from the MPEG-2 and MPEG-4 etenderde.

FIG. 2 illuetretes the presentetion process in the terminel erchitecture of FIG. 1 in eccordence with the present invention.

From the list of objects 126, the presentation engine 150 obtains e list of displeysbles (e.g., vidao objects) end audibles (e.g., audio objects). The liet of displeysbles end audibles is created and maintained by the composition engine 120, ee diecuseed.

The presentetion engine 150 elso renders the objecte to be preeented into the eppropriate freme buffers. The dieplayable objects ara rendered into the dieplay buffer 160, while the audible objects are rendered into the audio buffer 170. For this purpoee, the presentetion engine 150 intersects with the lowar level rendering libreries disclosed in the MPEG-4 standard.

The presentation engine 150 converts the content in the composition buffers 126, . . . , 136 into the eppropriete format before baing rendered into the display or audio buffere 160, 170 for

presentation on a display 240 and audio pleyer 242, respectively.

The presentation engine 150 is also responsible for efficient rendering of presentable content including rendering optimization, scelebility of the rendered data, end so forth.

Accordingly, it can be seen that the preeent invention providas a method and apparatus for compoeing and presanting multimedia programs using the MPEG-4 standerd. A multimedie terminal includas a terminal manager, e composition engine, content decoders, end e presentetion engine. The composition engine mainteins end updetes e scens greph of the currant objects, including their positions in a ecane and thair charactarietice, to provide e list of objects to be displayed to the presentetion engine. The presentation engine retrievee the corresponding objecte from content decoder buffers eccording to time stamp information.

The presentetion engine aesembles the decoded objects according to the list to provide e ecene for display on display devices, such ee e video monitor end speekers, end/or for storege on e etorege device.

The terminal manager receives ueer commands end causes the composition engine to update the scena greph end list of objects in response thereto. The terminal manager elso forwards object descriptore to e scene decoder at the composition engine.

Moreover, the composition engine and the presantetion engine preferebly run on eeperate

control threads. Appropriate interfece definitions can be provided to ellow the composition engine end the presentation engine to communicate with each othsr. Such interfeces, which can be developed using techniques known to those skilled in the art, should ellow the passing of messeges and data between the presentetion engine and the composition engine.

Although the invention has been described in connection with various specific embodiments, those skilled in the ert will eppreciets that numerous adaptetions and modifications may be mads thereto without deperting from the spirit end scope of the invention as set forth in the claims.

For example, while verious syntax elements have been diecussed herein, note that they sre exemples only, end any syntax may be used.

Moreover, while the invention hes been discussed in connection with the MPEG-4 stendard, it should be appreciated thet the concepts disclosed hsrein can be adapted for use with any similar communicetion stsnderds, including derivations of the current MPEG-4 stenderd.

Furthermore, the invention is suitable for use with virtually eny type of network, including cable or setellite televieion broedbsnd communication nstworks, locel srse networke (LANs), metropoliten area networks (MANs), wide area networks (WANs), internets, intranets, end the Internet, or combinations thereof.

What is claimed is:

1. A terminal for receiving and processing a multimedia data bitstrsem, comprising:

a terminal manager;

e composition engine;

e plurality of content decoders; end

s presentation engine; wherein:

said content decoders recover and decode multimedia objects from respective elementary streams of ths bitstreem;

ssid multimedie objscts comprising at leest one of video objects snd audio objecte for presentation in s multimedis scene;

said composition engine recovere scene description information from ths bitstresm that defines specific ones of the recovered multimedis objects that ere to be provided in the multimedie ecene, and characteristics of the recovered multimedie objects in the multimedia scene;

said terminal manager recovers object descriptor information from the bitstreem that associetes ssid recovered multimedis objects with respective ones of said elementery streams, and providee the recovered object deecriptor information to said composition engine;

seid composition engine is responsive to said recovered object descriptor information provided thereto end said recovered scene description information for creating e liet of said specific

ones of tha recovered multimedia objects that are to be displayed in said multimedia scens; snd

said presentation engine obtains said list from said composition engine, snd, in responee thereto, retrieves the corresponding decoded multimedia objecte from said content decodere to provide data corresponding to the multimedia scene to sn output device.

2. The terminsl of claim 1, wherein:

said compoeition engine snd ssid presentation engine have separste control thresds.

3. The terminsl of claim 2, wherein:

said saparate control thraads allow the presentation engine to begin retrieving the corresponding decoded multimedia objects while tha composition engina recovers additional scens description information from the bitstresm and/or prooesses additional object descriptor information provided thereto.

4. The terminal of claim 1, wherein:

said content decoders, presentation engine and composition engine hsve separate control threads.

5. The terminal of claim 1, wherein:

ssid chsracteristics of the racovered multimedia objects in the multimedia scene include positions of said specific onee of the recovered multimedis objects in said multimedis scene.

6. The terminsl of claim 1, wherein:

said recovered scene description information is provided according to a Binary Format for Scenee (BIFS) language

7. The terminal of claim 1, wherein:

said multimedia data bitstreem is provided according to sn MPEG-4 etandsrd.

8. The terminal of claim 1, wherein:

said compoeition engine maintsins scene grsph information of a compoeition of ssid multimedia scene in response to ssid recovered object deecriptor information provided thereto and said recovered ecene description information for use in creating said list.

9. The tarminal of claim 8, wherein:

said composition engine updates the ecene graph information, and eaid liet, as required, for successive multimedia scenes in response to subsequent recovered scene description information from the bitstresm.

10. The terminal of claim 8, wherein:

ssid terminal mansger is responsive to user input events at a user interfsce for providing corresponding dsta to said composition sngine for modifying ssid scene grsph, and said list, as required.

11. The terminal of claim 1, wherein:

said composition engine provides said list to said presentation engine according to a specified presentation rate.

12. The terminal of claim 1, wherein said multimedia objects comprise video and audio objects for presentation in the multimedia scene, further comprising:

video and audio buffers for buffering the video and audio objects, respectively, prior to presentation;

wherein said presentation engine reads objects from said list and provides them to the appropriate one of said video and audio buffers.

13. A terminal for receiving and processing a multimedia data bitstream, comprising:

decoding means for recovering and decoding multimedia objects from respective elementary streams of the bitstream;

said multimedia objects comprising at least one of video objects and audio objects for presentation in a multimedia scene;

composing means for recovering scene description information from the bitstream that defines specific ones of the recovered multimedia objects that are to be provided in the multimedia scene, and characteristics of the recovered multimedia objects in the multimedia scene;

managing means for recovering object descriptor information from the bitstream that associates said recovered multimedia objects with respective ones of said elementary streams, and providing the recovered object descriptor information to said composing means;

said composing means being responsive to said recovered object descriptor information provided thereto and said recovered scene description information for creating a list of said specific ones of the recovered multimedia objects that are to be displayed in said multimedia scene; and

presenting means for obtaining said list from said composing means, and, in response thereto, retrieving the corresponding decoded multimedia objects from said decoding means to provide data corresponding to the multimedia scene to an output device.

14. A method for receiving and processing a multimedia data bitstream at a terminal, comprising the steps of:

recovering and decoding multimedia objects from respective elementary streams of the bitstream at respective content decoders;

said multimedia objects comprising at least one of video and audio objects for presentation in a multimedia scene;

recovering scene description information from the bitstream that defines specific ones of the recovered multimedia objects that are to be provided

in the multimedia scene, and characteristics of the recovered multimedia objects in the multimedia scene;

recovering object descriptor information from the bitstream that associates said recovered multimedia objects with respective ones of said elementary streams;

creating a list of said specific ones of the recovered multimedia objects that are to be displayed in said multimedia scene in response to said recovered object descriptor information and said recovered scene description information; and

retrieving the corresponding decoded multimedia objects in response to the list to provide data corresponding to the multimedia scene to an output device.

15. The method of claim 14, wherein:

said recovering steps are performed using control threads that are separate from said retrieving step.

16. The method claim 15, wherein:

said separate control threads allow the retrieving of the decoded multimedia objects to begin while the recovering of additional scene description information and/or the recovering of additional object descriptor information occurs.

17. The method of claim 14, wherein:

said creating step is performed using a control thread that is separate from said retrieving step.

18. The method of claim 14, wherein:

said recovering steps and said creating step are performed using control threads that are separate from said retrieving step.

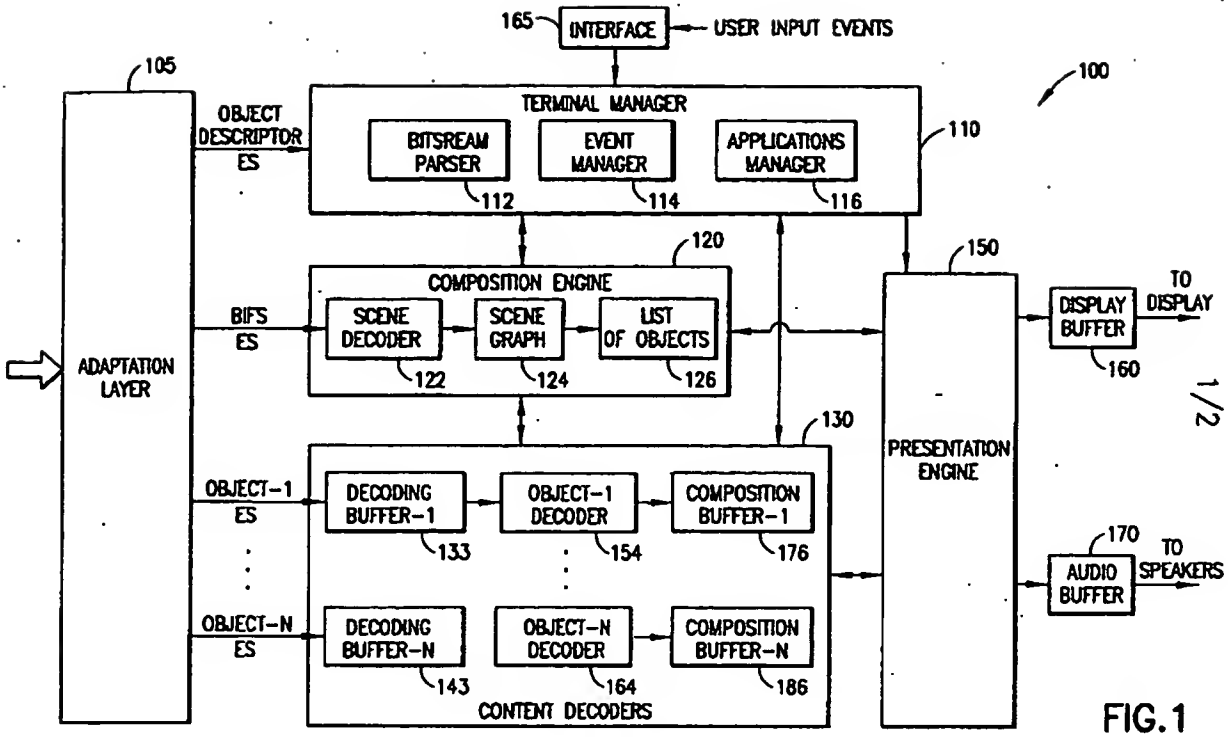


FIG.1

1/2

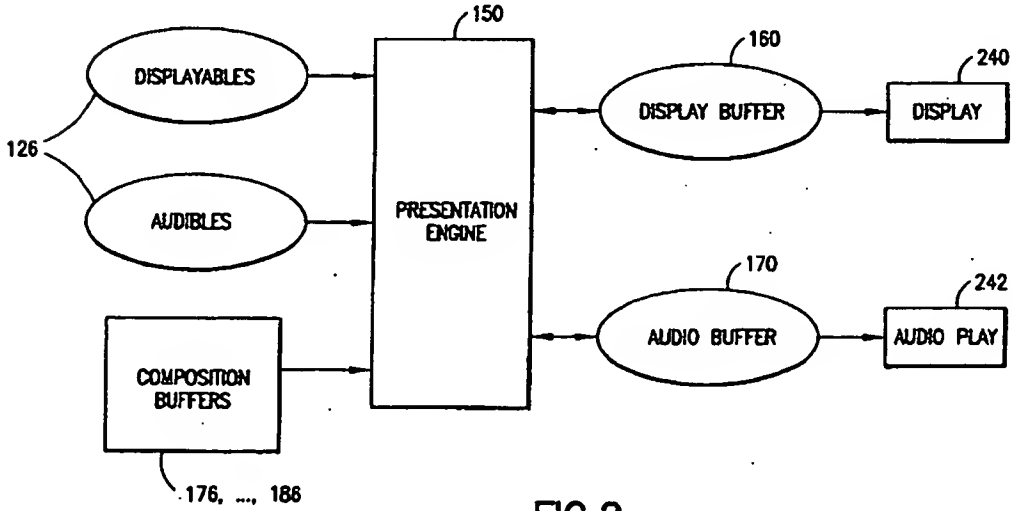


FIG.2

2/2

WO 00/01154 PCT/US99/14306

INTERNATIONAL SEARCH REPORT

International Application No
PC./US 99/14306

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04N7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CASALINO F ET AL: "MPEG-4 systems, concepts and implementation" MULTIMEDIA APPLICATIONS, SERVICES AND TECHNIQUES - ECMAST'98. THIRD EUROPEAN CONFERENCE. PROCEEDINGS, MULTIMEDIA APPLICATIONS, SERVICES AND TECHNIQUES - ECMAST '98 THIRD EUROPEAN CONFERENCE PROCEEDINGS, BERLIN, GERMANY, 26-28 MAY 1998, pages 504-517, XP002120837 1998, Berlin, Germany, Springer-Verlag, Germany ISBN: 3-540-64594-2 paragraph '0005! figure 3 --- -/--	1-18

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 October 1999	12/11/1999
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Berbain, F

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PC./US 99/14306

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ELEFTHERIAOIS A: "The MPEG-4 system and description languages: from practice to theory" PROCEEDINGS OF 1997 IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS. CIRCUITS AND SYSTEMS IN THE INFORMATION AGE. ISCAS '97 (CAT. NO.97CH35987), PROCEEDINGS OF 1997 IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS. CIRCUITS AND SYSTEMS IN TH, pages 1480-1483 vol.2, XP002120838 1997, New York, NY, USA, IEEE, USA ISBN: 0-7803-3583-X paragraphs '0002!,'0003! figure 1 ---	1-18
A	AVARO O ET AL: "The MPEG-4 systems and description languages: A way ahead in audio visual information representation" SIGNAL PROCESSING. IMAGE COMMUNICATION, vol. 9, no. 4, 1 May 1997 (1997-05-01), page 385-431 XP004075337 ISSN: 0923-5965 paragraph '0002! figure 5 ---	1-18.

Form PCT/ISA/210 (continuation of second sheet) (July 1992)